# Cloning and sequence analysis of *arom* gene from *Sclerotinia sclerotiorum*

YANG Qian[1], YU Han-ying[1,2], LI Chang-yin[1]

[1] *Department of Life Science and Engineering, Harbin Institute of Technology, Harbin 150001, P. R. China*

[2] *Department of Petroleum Engineering, Daqing Petroleum Institute, Daqing 163318, P. R. China*

**Abstract:** An *arom* gene was cloned from genomic DNA of *Scleortinia sclerotiorum* by inverse PCR. The evolutionary relationships of *S. sclerotiorum* and other fungi in *arom* gene were studied. Results showed that the *arom* gene from of *S. sclerotiorum* has a single open reading frame of 4 773 bp and does not include any introns. The derived amino acid sequence consists of 1 590 residues, and it is homologous to all fungal AROM proteins studied so far. The theoretical isoelectric point (pI) and molecular weight (Mw) is 6.5 and 172.66 kD, respectively. GC percentage of the *arom* gene is 44.94. According to the results of searching from CDD and Prosite database, AROM protein of *S. sclerotiorum* contains five conserve domains: 3-dehydroquinate synthase domain, 3-dehydroquinate dehydratase (3-dehydroquinase) domain, shikimate 5-dehydrogenase domain, shikimate kinase domain, and -enolpyruvylshikimate-3-phosphate synthase (EPSP sythase) domain, and four motifs: two EPSP synthase signatures, dehydroquinase class I active site, shikimate kinase signature. According to the PIR Site Rule PIRSR000514-1, four functionally important amino acid residues are found by alignment. Putative TATA box and CAAT box locate separately in -23 and -77 loci in 5′ un-translated region, and two loci found in downstream *arom* gene are likely polyadenylation signals. In addition, phylogeny of *arom* gene is analyzed.

**Keywords:** *Sclerotinia sclerotiorum*; *arom*; Gene cloning

**CLC number:** Q781 **Document code:** A **Article ID:** 1007-662X(2005)04-0260-05

## Introduction

*Sclerotinia sclerotiorum* is a phytopathogenic fungus, which can infect about 450 kinds of plants including many crops and leading to serious economic loss to agriculture production. The study on the fungus mainly focused on the control of sclerotinia diseases, the infection mechanisms, and its applications as a bio-herbicide and a source of polysaccharide. AROM is a pentafunctional protein which catalyzes five consecutive steps of aromatic amino acid biosynthetic pathway in fungi. Since the pathway is absent in mammal, the corresponding enzymes become the interesting targets to control the bacterial and fungal pathogens. In contrast to fungi, these metabolic steps in bacteria are catalyzed by monofunctional enzymes and the encoding genes locate separately in genome. Three separate peptides and one bifunctional protein containing the 3-dehydroquinate (DHQ) dehydratase and 5-dehydroshikimate (DHS) dehydrogenase activities exist in plants. The fungal arom locus is thought to have arisen through gene fusion because the protein is homologous with the five monofunctional bacterial enzymes and is made up of functional domains which fold independently or semi-independently (Banerji *et al.* 1993). Usually there is no intron or just a short one at the 3′ -terminal in the *arom* gene. Originally *arom* gene was considered to be a possible operon in eukaryotes similar to that in prokaryotes. However it has been confirmed to encode a peptide (Giles *et al.* 1967). Another reason that it is studied nowadays is that its 5-enolpyruvylshikimate 3-phosphate synthase (EPSP synthase) domain is the target of environment friendly herbicide glyphosate (McDowell *et al.* 1996). The *arom* genes of *Aspergillus nidulans, Saccharomyces cerevisiae, Pneumocystis carinii,* etc. have been cloned (Banerji *et al.* 1993, Charles *et al.* 1986, Duncan *et al.* 1987). Therefore, it is possible to study on the phylogeny using the constitutive gene. This paper describes the cloning and characterization of the *arom* gene from *S. sclerotiorum.*

**Biography:** YANG Qian (1959- ), male, Ph. D., Professor of Harbin Institute of Technology, Harbin 150001, P. R. China. E-mail: yangq@hit.edu.cn

**Received date:** 2005-06-18

**Accepted date:** 2005-07-20

**Responsible editor:** Song Funan

## Materials and methods

### Materials

*Escherichia coli* DH5α and *S. sclerotiorum* were preserved in microbiology laboratory of Life Science and Engineering Department in Harbin Institute of Technology. *Ex Taq*™ and pMD18-T were from Takara Biotechnology (Dalian) Co., Ltd.. T4 DNA ligase was from Promega Biosciences, Inc.. Degenerate and specific primers used in PCR were synthesized by Shanghai Sangon Biological Engineering Technology and Service CO., LTD. and SBS Genetech Co., Ltd. DNA Gel Extraction Kit was from V-gene Biotechnology Limited. DIG Random Labeling and Detection Kit II (AP) were from Wuhan Boster Biological Technology Ltd.. GeneRuler™ 1kb DNA Ladder was from Fermentas International Inc.

### DNA extraction

*S. sclerotiorum* was cultured in PD medium for DNA extraction (Sambrook and Russell 2002).

### Amplification and cloning of an *arom* gene fragment

An *arom* gene fragment was cloned by a pair of degenerate primers. The primers were designed according to the alignment result of the known AROM protein sequences from *Neurospora crassa, Emericella nidulans, Aspergillus fumigatus* and *Thanatephorus cucumeris* by ClustalW. The sense primer was 5′ GCT TCY CGT TTC CTY ACY ACY GTR 3′ and the antisense primer was 5′ TTC GCY MGY TTC CTI CGT RYY ATY 3′. PCR was performed at 94°C for 30 s, at 40°C for 30 s and 72°C for 2 min for 4 cycles, then at 94°C for 30 s, at 60°C for 30 s and 72°C for 2 min for 31 cycles with *Ex Taq*™ enzyme. The PCR product was purified by gel extraction kit, and then linked to the

pMD18-T vector. Ligation solution consisted of: insert DNA 3 μL, vector 1 μL, $H_2O$ 4.7 μL, 10× ligase buffer 1 μL, T4 DNA ligase 0.3 μL. The constructed recombination plasmid was sequenced for the preparation of the Southern blot probe Ta and the design of specific primers used in inverse PCR.

**Southern analysis of the *arom* gene**

The probe Ta was labeled and Southern blotting was performed on *S. sclerotiorum* genomic digested DNA following the introduction of DIG Random Labeling and Detection Kit II (AP).

**Isolation and preparation of template DNA**

*S. sclerotiorum* genomic DNA were digested by the selected restriction enzymes *Sph* I, *Acc*II, *Afa* I, *Sph* I, *Taq* I, *Xba* I, *Xho* I, *HgiC* I, *Eco*T14 I, *Bgl*II, and *Hae*III for 20 h. *S. sclerotiorum* DNA digested by *Sph* I and *HgiC* I were collected according to the size of hybridization bands, and ligated by T4 DNA ligase.

**Amplification of 3'- and 5'-sequence of *arom* fragments by inverse PCR**

3'-teriminal sequence of the *arom* gene was cloned from *S. sclerotiorum* by inverse PCR. The 3'-terminal fragment was amplified with P1 and P2 following the introduction of *Ex Taq*™ in the presence of the regathered DNA intro-molecular ligation product (5 μL). PCR reaction cycle consisted of: 94 ℃ 30 s, anneal 30 s, 72 ℃ 6 min, 12 cycles (The anneal temperature of the first cycle was 57℃, then lowered 0.5℃ per cycle), and then 94 ℃ 30 s, 51 ℃ 30 s, 72℃ 6 min, 28 cycles. However, two inverse PCR were done separately with the nested primers P1, P3 and P2, P4 in the touchdown condition in order to attain 5'-terminal sequence of the *arom* gene from *S. sclerotiorum*. The first PCR reaction cycle consisted of: 94 ℃ 30 s, anneal 30 s, 72℃ 4 min, 16 cycles (The anneal temperature of the first cycle was 57 ℃, then lowered 0.5 ℃ per cycle), and then 94 ℃ 30 s, 49℃ 30 s, 72℃ 4 min, 19 cycles. The fragment was amplified with the primer pairs P1 and P3 in the presence of 1 μL *HgiC* I digested genomic DNA intro-molecular ligation product. The second PCR reaction cycle consisted of: 94 ℃ 30 s, anneal 30 s, 72 ℃ 4 min, 12 cycles (The anneal temperature of the first cycle was 57 ℃, then lowered 0.5 ℃ per cycle) and 94 ℃ 30 s, 51 ℃ 30 s, 72 ℃ 4 min, 23 cycles. The amplification was carried out in the presence of the first diluted PCR product (1 μL) with P2 and P4.

P1: 5′ GTCAAGAAAGCATCCGTC 3′;
P2: 5′ CGAATGTGTTGGGAAGAC 3′;
P3: 5′ CAAAGCAATGAAGGATGA 3′;
P4: 5′ ATACGGTGGCTGCTAAAT 3′.

The primers used to identify the presence of the fragment of *arom* gene were separately P3, P5 and P6, P7.

P5: 5′ GGAATGACTTGGAGAGGA 3′;
P6: 5′ GCTCAGCGACTTATCCAC 3′;
P7: 5′ CAAACACATGCTTCTCTG 3′.

**DNA sequencing**

DNA sequencing was performed by Shanghai Sangon Biological Engineering Technology and Service Co., Ltd. and Shanghai Boya Biological Technology Co., Ltd.

## Results

**Cloning of the *arom* gene from *S. sclerotiorum***

A fragment (1 626 bp) containing shikimate kinase domain and partial EPSPs domain was obtained from *S. sclerotiorum* genomic DNA with the degenerate primers and was linked to the pMD18-T vector (Fig. 1). Southern blot proved that it originated from the fungus (Fig. 2) and showed the fragments possibly included the *arom* gene segment were about 6 000 bp, 4 500 bp and 2 700 bp in length in the *Sph* I digested genomic DNA. Subsequently, the DNA fragments with sizes between 6 000 and 8 000 bp, 4 000 and 5 000 bp as well as 2 500 and 3 000 bp were excised from the agarose gel and used for the intro-molecular ligation. The intro-molecular ligation product served as the template of inverse PCR and the 3'-terminal sequence of the *arom* gene from *S. scleortiorum* was obtained (Fig. 3). Another digestion by *HgiC* I, *Bgl*II and *Eco*T14 I were prepared for Southern blot and the hybridization bands about 4 000 bp, 7 000 bp and 2 000 bp in size were shown (Fig. 2). The 5'-terminal sequence of the *arom* gene from *S. scleortiorum* was cloned with the complete *HgiC* I digested genomic DNA as the template (Fig. 3). The *arom* gene from *S. scleortiorum* has been submitted to GenBank and the accession number is AY746008.

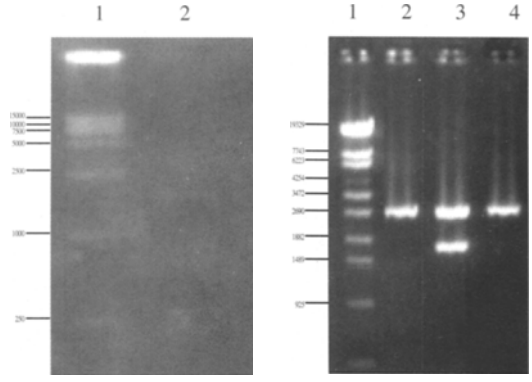


**Fig. 1 A: Amplification of a fragment of the *arom* gene by PCR.** 1. DNA marker; 2. PCR fragment. **B: Analysis of recombined plasmid by digestion.** 1. DNA marker; 2, 3, 4. Digestion fragments.

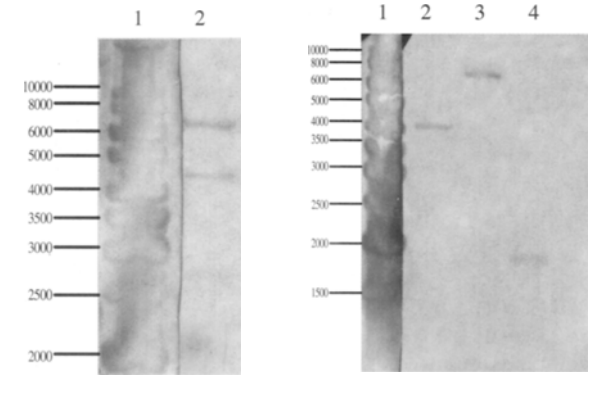


**Fig. 2 A: Genomic southern blot analysis I of *Scleortinia sclerotiorum*.** 1. Marker with itself as probe; 2. Genomic DNA/ *Sph* I with the cloned fragment of *arom* gene as probe. **B: Genomic southern blot analysis II of *Scleortinia sclerotiorum*.** 1. Marker with itself as probe; 2. Genomic DNA/ *HgiC* I; 3. Genomic DNA/ *Bgl*II; 4. Genomic DNA/*Eco*T14 I (2,3,4 with the cloned fragment of the *arom* gene as probe).

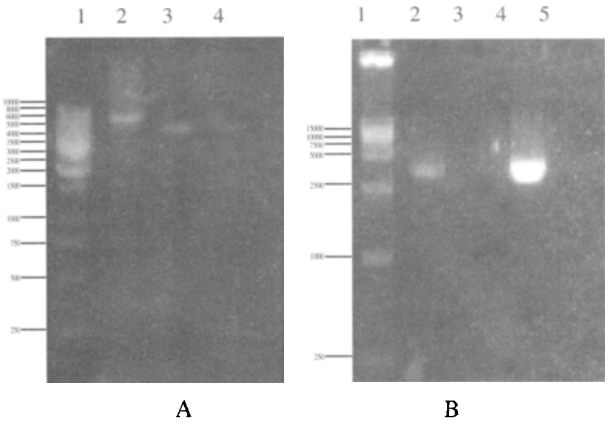


**Fig. 3 A: Amplification of 3′-terminal sequence of *arom* gene by PCR.** 1. DNA marker; 2. PCR fragment using 6000 to 8000bp of DNA as the template; 3. PCR fragment using 4000 to 5000bp of DNA as the template; 4. PCR fragment using 2500 to 3000bp of DNA as the template.

**Fig. 3B: Amplification of 5′-terminal sequence of *arom* gene by PCR.** 1. DNA marker; 2. Nested PCR fragment using the first PCR product diluted 10 times as the template; 3. Nested PCR fragment using the first PCR product diluted $10^2$ times as the template; 4. Nested PCR fragment using the first PCR product diluted $10^3$ times as the template; 5. Nested PCR fragment using the first PCR product diluted $10^4$ times as the template.

### Sequence analysis of *S. sclerotiorum arom* gene and AROM protein

The *arom* gene of *S. sclerotiorum* is homologous to those of *Saccharomyces cerevisiae* (X06077) and *Aspergillus nidulans*(X05204). G+C percentage of the *arom* gene is 44.94%. There is a single open reading frame of 4 773 bp determined by ORF Finder and analogy. The derived amino acid sequence from the ORF consists of 1 590 residues and is 76% homologous to the AROM protein of *Aspergillus fumigatus* (CAD29607), 65% to *Pneumocystis carinii* (Q12659), etc. Although the BLASTX alignment shows that several amino acids at the C-terminal of stop codon are matched to the known AROM protein, there is no GT-AG or lariat forming motif downstream of the putative ORF and two other stop codons just behind the opal codon are found in the same ORF. Estimated molecular weight of the putative AROM protein is 172 . 66 kD and pI is 6.45. The possible promoter factor such as TATA box, CAAT box, and GCN4 binding site are found at proper nucleotide position (TATA box at nucleotide position -23, CAAT box at -77, GCN4 binding site at -173). However, two possible polyadenylation signal motifs are found at 3′-untranslated region and can not been determined, of them one is correct up to date. They are shown in Fig.4 with initiator, initiation codon, and stop codon (Opal codon).

The locations of the domains in the *S. sclerotiorum* protein, corresponding to the five enzyme activities, are determined by searching CDD database (Marchler-Bauer *et al.* 2003). The order of the functional domains from the N- to the C-terminus are DHA synthase (30-393), EPSPS (405-835), shikimate kinase (864-1035), DHQ dehydratase (1052-1279), and shikimate dehydrogenase (1308-1557) (Fig. 4). Throughout the *S. sclerotiorum* AROM protein there are four important amino acid motifs which are conserved in equivalent enzymes across fungal, bacterial, and plant species: EPSP synthase signature 1, LyiGNAGTAsRfLtT (486–500); EPSP synthase signature 2, RvKECNRIkAMkdeLakFG (748–766); Shikimate kinase signature, RaaEldlLkhcltdqpekh. VFAcGGGvV (919–946); Dehydroquinase class I active site, DLEIafpdellqevteakgfsr. IIa SHHD (115 2–1180) (Fig. 4). Four functionally important amino acid residues (K155, K253, R267, H290) in *S. sclerotiorum* AROM protein are determined by the alignment and PIR Site Rule PIRSR000514-1 (K152, K250, R264, H287) (Fig. 4). In addition, phylogenetic analyses of several fungi are performed based on distances from AROM protein sequences (Fig. 5).

## Discussion and conclusion

AROM is a key enzyme in biosynthetic pathway of aromatic amino acid in fungi, which provides an economical means of simultaneously synthesizing five shikimate pathway enzymes. In this study, the *arom* gene of *S. sclerotiorum* was cloned by inverse PCR. It is known that AROM protein encoded by *arom* gene locates in cytoplasm and there is no signal peptide. Minimum, maximum and average length of the pentafunctional AROM protein known are 1 563, 1 618 and 1 585 amino acids, respectively. AROM protein contains two EPSP synthase signatures. The first pattern (486–500) corresponds to a region that is a part of the active site and which is also important for the resistance to glyphosate. The second pattern (919–946) is located in the C-terminal part of EPSP synthase and contains a conserved lysine which seems to be important for the activity of the enzyme (Padgette *et al.* 1991). Two classes of dehydroquinases exist, known as types I and II. The 3-dehydroquinate dehydratase domain of AROM protein has the dehydroquinase class I active site signature. The amino acid H in the signature is the active site residue (Deka *et al.* 1992). There is one shikimate kinase signature in AROM protein sequence. The motif is a glycine-rich region, which typically forms a flexible loop between a beta-strand and an alpha-helix. This loop interacts with one of the phosphate groups of the nucleotide and forms the ATP/GTP-binding site motif A (P-loop) (Walker *et al.* 1982; Saraste *et al.* 1990).

The result of genomic Southern blot analysis II of *S. sclerotiorum* indicated that probably only one copy of the *arom* gene exists in *S. sclerotiorum*. How many copies of the *arom* gene in *S. sclerotiorum* could not be determined by the result of genomic southern blot analysis I although it showed three bands in each lane. The reason is that: the inverse PCR fragments about 4 000 bp and 3 000 bp in size (lane 3 and lane 4 in Fig.3 A) were not sequenced; the hybridization bands about 4 000 bp and 3 000 bp in size perhaps were not the segments of *arom* gene from *S. sclerotiorum*, but were resulted from unspecific hybridization, incomplete *Sph* I digestion or mechanical split of *S. sclerotiorum* genomic DNA.

The *arom* gene locus of *S. sclerotiorum* seemed to possess a GCN4 binding site. The optimum sequence for this motif, usually located at upstream (100-600 bp) of the transcriptional start site, is the palindrome ATGA(C/G)TCAT, but the sequence (TTGAGTCAC) found upstream of the *arom* locus of *S. sclerotiorum* shows a conservation of the essential bases. The binding of GCN4 to its target sites mediates the "general amino acid control" mechanism of *S. cerevisiae* in which starvation for a particular amino acid can lead to the up-regulation of expression of more than 30 genes involved in amino acid biosynthesis (Fink 1986). The ARO1 (AROM encoding) locus of *S. cerevisiae* has been shown to be under general amino acid control (Duncan *et al.* 1988).

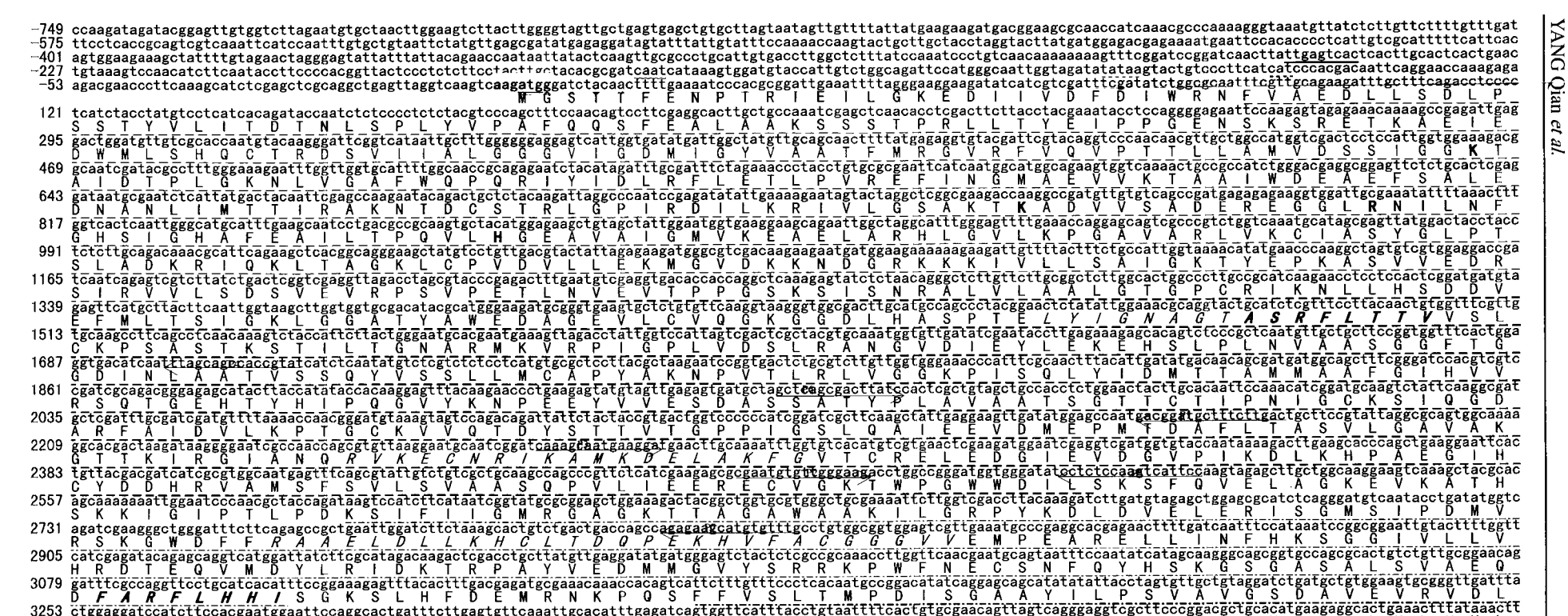

**Fig. 4 The analysis of *arom* gene of *S. scleortiorum.***

TATA box; Initiator; Initiation codon; Kozak; CAAT box; GCN4 binding site; * stop codon; possible polyadenylation signal; 3-dehydroquinate synthase domain, 5-enolpyruvylshikimate-3-phosphate synthase domain, Shikimate kinase domain, 3-dehydroquinate dehydratase domain, Shikimate 5-dehydrogenase domain; The italic amino acids were two EPSP synthase signature II, shikimate kinase signature and dehydroquinase class I active site signature respectively; The boldfaces were functionally important amino acids; The italic boldfaces were the amimo acids used for degenerate primers designing. Other primers used were indicated by the arrows. P1, P4, P5 and P7 were represented by the complemented sequences

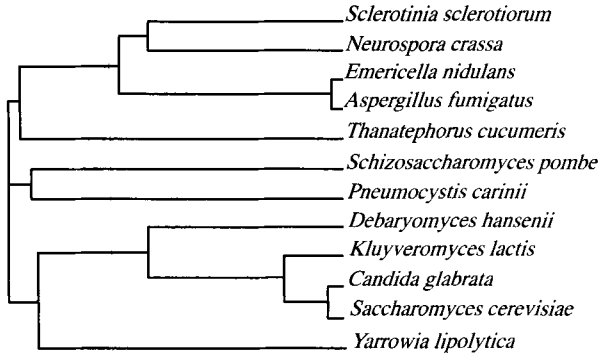


**Fig. 5 Phylogenetic analysis of several fungi based on AROM protein**

The alignment of the deduced amino acid sequence of the *arom* gene from *S. sclerotiorum* by BLASTP also included the amino acid sequences of the *qut* repressors from *N. crassa* and *A. nidulans*, which were homologous to the three domains at the C-terminal of the AROM proteins. This relationship is interesting in both functional and evolutionary terms. The repressor is thought to be capable of binding quinate, DHQ and DHS such that the binding of any of these would lead to the inactivation of the *qutR* protein. However, in the absence of these metabolites the activator of the quinate utilization pathway can't bring about the induction of the enzymes required for quinate catabolism. The kinase activity of the *qut* repressor is thought to be involved in the inactivation of the activator (Hawkins *et al.* 1991). The homology between the *qut* repressor and AROM suggests an evolutionary link between the two genes.

In conclusion, the homology between the deduced amino acid sequence from the putative *arom* gene of *S. sclerotiorum* and the other AROM protein sequences together with the conservation of motifs identified as functionally important suggested very strongly that the gene encoded a functional AROM protein *in vivo*. However, formal proof that the gene encodes a functional protein awaits the expression of the cloned gene in a heterologous system such as *Trichdoma viride* or *E. coli* and enzyme assay.

## References

Banerji, S., Wakefield, A.E., Allen, A.G., Maskell, D.J., Peters, S.E., Hopkin, J.M. 1993. The cloning and characterization of the *arom* gene of *Pneumocystis carinii* [J]. J Gen Microbiol, **139**(12): 2901–2914.

Charles, I.G., Keyte, J.W., Brammar, W.J., Smith, M., Hawkins, A.R. 1986. The isolation and nucleotide sequence of the complex AROM locus of *Aspergilus nidulans* [J]. Nucleic Acids Research, **14**(5): 2201–2213.

Deka, R.K., Kleanthous, C., Coggins, J.R. 1992. Identification of the essential histidine residue at the active site of *Escherichia coli* dehydroquinase [J]. J Biol Chem, **267**(31): 22237–22242.

Duncan, K., Edwards, R.M., Coggins, J.R. 1987. The pentafunctional arom enzyme of *Saccharomyces cerevisiae* is a mosaic of monofunctional domains [J]. Biochem J, **246**(2): 375–386.

Duncan, K., Edwards, R.M., Coggins, J.R. 1988. The *Saccharomyces cerevisiae* ARO1 gene: an example of the co-ordinate regulation of five enzymes on a single pathway [J]. FEBS Letters, **241**(1–2): 83–88.

Fink, G.R. 1986. Translational control of transcription in eukaryotes [J]. Cell, **45**(2): 155-156.

Giles, N.H., Case, M.E., Partridge, C.W.H., Ahmed, S.I. 1967. A gene cluster in *Neurospora crassa* coding for an aggregate of five aromatic synthetic enzymes [J]. Genetics, **58**: 1453–1460

Hawkins, A.R., Lamb, H.K., Roberts, C.F. 1991. Structure of the *Aspergillus nidulans qut* repressor-encoding gene: implications for the regulation of transcription initiation [J]. Gene, **110**(1): 109–114.

Marchler-Bauer, A., Anderson, J.B., DeWeese-Scott, C., Fedorova, N.D., Geer, L.Y., He, S., Hurwitz, D.I., Jackson, J.D., Jacobs, A.R., Lanczycki, C.J., Liebert, C.A., Liu, C., Madej, T., Marchler, G.H., Mazumder, R., Nikolskaya, A.N., Panchenko, A.R., Rao, B.S., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Vasudevan, S., Wang, Y., Yamashita, R.A., Yin, J.J., Bryant, S.H. 2003. CDD: a curated Entrez database of conserved domain alignments [J]. Nucleic Acids Research, **31**(1): 383–387.

McDowell, L.M., Schmidt, A., Cohen, E.R., Studelska, D.R., Schaefer, J. 1996. Structural constraints on the ternary complex of 5-enolpyruvylshikimate-3-phosphate synthase from rotational-echo double-resonance NMR [J]. J Mol Biol, **256**:160–171.

Padgette, S.R., Re, D.B., Gaser, C.S., Eichholtz, D.A., Frazier, R.B., Hironaka, C.M., Levine, E.B., Shah, D.M., Fraley, R.T., Kishore, G.M. 1991. Site-directed mutagenesis of a conserved region of the 5-enolpyruvylshikimate-3-phosphate synthase active site [J]. J Biol Chem, **266**(33): 22364–22369.

Walker, J.E., Saraste, M., Runswick, M.J., Gay, N.J. 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold [J]. EMBO J, **1**(8): 945–951.

Sambrook, J., Russell, D.W. 2002. Molecular Cloning: A laboratory Manual, 3rd Ed [M]. New York: Cold Spring Harbor Laboratory Press.

Saraste, M., Sibbald, P.R., Wittinghofer, A. 1990. The P-loop--a common motif in ATP- and GTP-binding proteins [J]. Trends Biochem Sci, **15**: 430–434.